

```
library(tidyverse)
library(stringr)

# Download the txt file GSE155153_ZTOAll_step2_washU_text.txt.gz from https://www.ncbi.nlm.nih.gov/geo/
input <- read.delim("data/Promoter_Capture_HiC/GSE155153_ZTOAll_step2_washU_text_DOWNLOADED_FROM_GEO.txt")
```

To be uploaded at UCSC

```
# Modify the file for UCSC assembly tool conversion.
input2 <- input %>% separate(col = V1, into = c("chrom", "chromStart", "chromEnd"), sep = ",") %>%
  mutate("name" = paste0("ident_", row_number()), .before=V2) %>%
  mutate("score" = round(V3)) %>%
  mutate("value" = "0") %>%
  mutate("exp" = "exp") %>%
  mutate("color" = "#cccccc") %>%
  mutate("sourceChrom" = chrom) %>%
  mutate("sourceStart" = chromStart) %>%
  mutate("sourceEnd" = chromEnd) %>%
  mutate("sourceName" = paste0("ident_", row_number())) %>%
  mutate("sourceStrand" = ".") %>%
  separate(col = V2, into = c("targetChrom", "targetStart", "targetEnd"), sep = ",") %>%
  mutate("targetName" = paste0("ident_", row_number())) %>% # We need an identifier here
  mutate("targetStrand" = ".") %>%
  dplyr::select(chrom, chromStart, chromEnd, name, score, value, exp, color, sourceChrom, targetChrom, targetStart, targetEnd, targetStrand)

#write.table(input2, "GSE155153_ZTOAll_step2_UCSC1.txt", quote=F, row.names = F, sep="\t")

# Have to lift over all the coordinates first. --> need identifier.
input_source <- input2 %>% dplyr::select(1:5)
#write.table(input_source, "6_CaptureHiC_GSE155153/GSE155153_ZTOAll_step2_UCSC_sourcecm9.txt", quote=F, row.names = F, sep="\t")

input_target <- input2 %>% dplyr::select(14:18)
#write.table(input_target, "6_CaptureHiC_GSE155153/GSE155153_ZTOAll_step2_UCSC_targetmm9.txt", quote=F, row.names = F, sep="\t")

# Try a format like this: #chrom chromStart chromEnd name score value exp color sourceChrom sourceStart sourceEnd sourceName sourceStrand targetChrom targetStart targetEnd targetStrand
head(input2)
```

```
##   chrom chromStart chromEnd   name score value exp  color sourceChrom
## 1  chr1    3666698  3669330 ident_1     8     0 exp #cccccc      chr1
## 2  chr1    3666698  3669330 ident_2     6     0 exp #cccccc      chr1
## 3  chr1    4350079  4352609 ident_3     7     0 exp #cccccc      chr1
## 4  chr1    4485600  4488420 ident_4     6     0 exp #cccccc      chr1
## 5  chr1    4485600  4488420 ident_5     5     0 exp #cccccc      chr1
## 6  chr1    4485600  4488420 ident_6     7     0 exp #cccccc      chr1
##   sourceStart sourceEnd sourceName sourceStrand targetChrom targetStart
## 1      3666698   3669330   ident_1           .        chr1      3455951
## 2      3666698   3669330   ident_2           .        chr1      3659706
## 3      4350079   4352609   ident_3           .        chr1      4733719
## 4      4485600   4488420   ident_4           .        chr1      4712573
```

```
## 5      4485600      4488420      ident_5      .      chr1      4744876
## 6      4485600      4488420      ident_6      .      chr1      4746796
##      targetEnd targetName targetStrand
## 1      3457756      ident_1      .
## 2      3665317      ident_2      .
## 3      4739024      ident_3      .
## 4      4713583      ident_4      .
## 5      4746795      ident_5      .
## 6      4753247      ident_6      .
```

USE THE UCSC ASSEMBLY TOOL @ <https://genome.ucsc.edu/cgi-bin/hgLiftOver> . Copy in the content of the two txt files, delete the header so the tool accepts the input. Download the resulting file, rename and import here. Might have to paste the headers back. Some sites may be lost due to conversionsm problems.

```
import_mm10_source <- read.delim("data/Promoter_Capture_HiC/GSE155153_ZT0All_step2_UCSC_assembly_tool_s
mutate(sourceChrom = paste0("chr", V1)) %>% dplyr::select(sourceChrom, sourceStart=V2, sourceEnd=V3, r

import_mm10_source <- import_mm10_source[!(duplicated(import_mm10_source$name)),]

import_mm10_target <- read.delim("data/Promoter_Capture_HiC/GSE155153_ZT0All_step2_UCSC_assembly_tool_t
mutate(targetChrom = paste0("chr", V1)) %>% dplyr::select(targetChrom, targetStart=V2, targetEnd=V3, r

length(unique(import_mm10_target$name))
```

```
## [1] 146261
```

```
import_mm10_target <- import_mm10_target[!(duplicated(import_mm10_target$name)),]
import_mm10 <- inner_join(import_mm10_source, import_mm10_target, by= "name")

import_mm10_2 <- import_mm10 %>%
  mutate(chrom = sourceChrom, .before = sourceChrom) %>%
  mutate(chromStart = sourceStart, .before = sourceChrom) %>%
  mutate(chromEnd = sourceEnd, .before = sourceChrom) %>%
  mutate("Name" = name, .before = sourceChrom) %>%
  mutate("score" = sourceStrand, .before = sourceChrom) %>%
  mutate("value" = "0", .before = sourceChrom) %>%
  mutate("exp" = "exp", .before = sourceChrom) %>%
  mutate("color" = "#000000", .before = sourceChrom) %>%
  mutate(targetName = name, .before = targetStrand)

import_mm10_2$sourceStrand <- "."
head(import_mm10_2)
```

```
##      chrom chromStart chromEnd      Name score value exp      color sourceChrom
## 1   chr1      3676617  3679249 ident_1      8      0 exp #000000      chr1
```

```

## 2 chr1 3676617 3679249 ident_2 6 0 exp #000000 chr1
## 3 chr1 4359998 4362528 ident_3 7 0 exp #000000 chr1
## 4 chr1 4495519 4498339 ident_4 6 0 exp #000000 chr1
## 5 chr1 4495519 4498339 ident_5 5 0 exp #000000 chr1
## 6 chr1 4495519 4498339 ident_6 7 0 exp #000000 chr1
## sourceStart sourceEnd name sourceStrand targetChrom targetStart targetEnd
## 1 3676617 3679249 ident_1 . chr1 3465870 3467675
## 2 3676617 3679249 ident_2 . chr1 3669625 3675236
## 3 4359998 4362528 ident_3 . chr1 4743638 4748943
## 4 4495519 4498339 ident_4 . chr1 4722492 4723502
## 5 4495519 4498339 ident_5 . chr1 4754795 4756714
## 6 4495519 4498339 ident_6 . chr1 4756715 4763166
## targetName targetStrand
## 1 ident_1 .
## 2 ident_2 .
## 3 ident_3 .
## 4 ident_4 .
## 5 ident_5 .
## 6 ident_6 .

```

```

#write.table(import_mm10_2, "data/Promoter_Capture_HiC/GSE155153_ZT0All_step2_UCSC_mm10_merged.txt", se

```